

ISBN
978-602-96343-1-0

PROSIDING
KONFERENSI ILMIAH NASIONAL
HIMPUNAN EVALUASI PENDIDIKAN INDONESIA

EVALUASI IMPLEMENTASI KURIKULUM 2013
DAN SISTEM PENILAIAN



MANADO, 20 – 21 SEPTEMBER 2013

HEPI UNIT KOORDINASI DAERAH MANADO

UNIVERSITAS NEGERI MANADO



DAFTAR ISI

	halaman
HALAMAN JUDUL	i
KATA PENGANTAR	ii
DAFTAR ISI	iii
A. Sistem Penilaian	
1. Estimasi Reliabilitas Multidimensi Menggunakan AnalisisFaktor Konfirmatori (Gaguk Margono)	1
2. Penentuan Kriteria Ketuntasan Minimal (KKM) Keterampilan Produktif Jurusan Bangunan di SMK (Djemari Mardapi, Sudji Munadi, Samsul Hadi, Heri Retnawati)	16
3. Implementasi Metode Nedelsky Untuk Penetapan Standar Kelulusan Ujian Nasional di Sekolah Dasar (Sri Rejeki)	32
4. Validitas dan Reliabilitas Instrumen Penilaian Seleksi Ujian Masuk Keterampilan Khusus Calon Mahasiswa Jurusan Pendidikan Seni Rupa FBS UNY (Trie Hartiti Retnowati, Bambang Prihadi, Mardiyatmo)	45
5. Pengembangan Instrumen Penilaian Kinerja Guru Praktikum Fisika SMA di Provinsi D. I. Yogyakarta. (Nurul Fitriyah Sulaeman, Badrun Kartowagiran)	59
6. Abilitas Komposit Dalam Tes Potensi (Saifuddin Azwar dan Ali Ridho)	74
7. Pengembangan Model Asesmen Autentik Dalam Pembelajaran IPA Kurikulum 2013 Sekolah Menengah Pertama (Undang Rosidin)	96

8. Kualitas Instrumen Penilaian Hasil Belajar Fisika Di SMAN 1 dan 2 Tondano (Aswin H. Mondolang)	109
9. Perbandingan Validitas Kriteria <i>Test Of English Proficiency</i> Terhadap <i>Itp-Toefl</i> (Estimasi Menggunakan Korelasi Biasa Dengan Korelasi Kanonis) (Heri Retnawati)	118
10. Analisis Validasi Soal Penilaian Hasil Belajar Pada Program Pembelajaran di Jurusan Pendidikan IPS FKIP Universitas Lampung (Edy Purnomo)	132
11. Karakteristik Penyetaraan Skor (<i>Equating</i>) Untuk Data Polytomus (Risky Setiawan)	146
12. Lembar Kegiatan Pembelajaran Berbasis Masalah Dalam Asesmen Alternatif Pada Mata Kuliah Pendidikan Konsumen Untuk Melatih dan Menilai Kemampuan Berpikir Kritis Mahasiswa (Sri Wening)	167
13. Meningkatkan Kualitas Kegiatan Praktikum Fisika Dasar Dengan Menggunakan Model Evaluasi Pada Mahasiswa FMIPA Unima (Anneke Tienneke Rondonuwu)	181
14. Differential Item Functioning Potensi Akademik Pada Kelompok SMA– MA (Ali Ridho)	192
15. Pengembangan Soal Pisa Matematika Pada Kompetensi Proses Koneksi dan Refleksi (Navel O. Mangelep)	205
16. Evaluasi Penyelenggaraan Ujian Nasional Tahun 2013 (Bambang Suryadi)	221
17. Evaluasi Program Implementasi Standar PAUD Studi Evaluatif di TK Negeri Pembina Kabupaten Ende Tahun Ajaran 2012/2013 (Maria Goreti V. Anamara, Myrnawati, C.H, Wardani Rahayu	234

DIFFERENTIAL ITEM FUNCTIONING POTENSI AKADEMIK PADA KELOMPOK SMA – MA¹

Ali Ridho²

ABSTRACT

Evidence of fairness indicated by item equality functioning between groups compared. Items unfairness showed by differential item functioning (DIF). This study aimed at identify items of Potensi Akademik (PA) Seleksi Penerimaan Mahasiswa Baru – Perguruan Tinggi Agama Islam Negeri (SPMB-PTAIN) that contained differential item functioning (DIF) based on graduated school (Madrasah Aliyah Negeri – Sekolah Menengah Atas Negeri, MAN – SMAN, MS) Analyzing to data response ($n = 10000$, 5000 from MAN and 5000 from SMAN) by likelihood ratio test (LRT), the result showed that: (1) 12 items of analogy, logic, arithmetic and comparison components were significantly contained DIF favored SMAN; (2) 3 items of analogy and arithmetic components contained DIF favored MAN.

Keyword: *item response theory (IRT), differential item functioning (DIF), potensi akademik, SPMB-PTAIN*

PENDAHULUAN

Melengkapi kebutuhan seleksi mahasiswa baru, sebagai pengelola Perguruan Tinggi Negeri (PTN) dengan karakteristik yang khas, Direktorat Pendidikan Tinggi Islam (Diktis) Kementerian Agama berinisiatif untuk mengembangkan sistem seleksi secara bersama bagi seluruh Perguruan Tinggi Agama Islam Negeri (PTAIN). Tahun 2009 hingga sekarang terealisasi sistem

¹ Tulisan ini adalah bagian dari Laporan Hibah Penelitian Fakultas Psikologi UIN Maliki Malang Tahun 2013, dipresentasikan pada Konferensi Ilmiah Himpunan Evaluasi Pendidikan (HEPI) di Manado, 20-21 September 2013.

² Dosen Psikometri di Fakultas Psikologi UIN Maliki Malang. Email: ali.ridho@live.com

Seleksi Penerimaan Mahasiswa Baru Perguruan Tinggi Agama Islam Negeri (SPMB-PTAIN). Sistem ini banyak mengadaptasi Seleksi Nasional Masuk Perguruan Tinggi Negeri (SNMPTN) yang sudah mendahului.

Ujian masuk perguruan tinggi (PT) adalah peristiwa yang krusial. Implikasinya begitu besar, berpengaruh terhadap masa depan para peserta ujian. Sebagai *high stakes testing*, alat ukur yang digunakan dalam ujian masuk PT seharusnya memiliki karakteristik psikometrik yang baik, dalam arti valid, reliabel, dan adil. Menurut Messick (1998), selain sebuah tes seharusnya didukung oleh bukti-bukti validitas dan reliabilitas, sebuah tes seharusnya juga didukung oleh bukti-bukti bahwa tes bersifat adil. Keadilan yang dimaksud oleh Messick mengacu pada kesetaraan interpretasi makna skor antar individu, kelompok, dan latar belakang peserta tes. Keadilan semacam ini dibuktikan dengan berfungsinya aitem-aitem tes secara setara antara kelompok-kelompok yang mengikuti tes. Apabila terjadi ketidaksetaraan fungsi aitem pada kelompok-kelompok yang berbeda, artinya aitem-aitem tersebut belum dapat disebut memiliki sifat adil. Ketidaksetaraan fungsi aitem seperti ini lazim disebut *differential item functioning* (DIF).

Peserta SPMB-PTAIN tidak hanya berasal dari lulusan Madrasah Aliyah (MA), namun terbuka juga bagi peserta lulusan Sekolah Menengah Atas (SMA), Sekolah Menengah Kejuruan (SMK), serta sekolah-sekolah lain yang sederajat. Mengingat bervariasinya asal sekolah peserta SPMB-PTAIN, agar dapat dikatakan bahwa keadilan telah tercapai, diperlukan bukti bahwa aitem-aitem tes PA telah berfungsi setara antar peserta yang berasal dari jenis sekolah yang berbeda. Bukti-bukti kesetaraan ini tercapai manakala aitem-aitem dalam PA terbebas dari DIF. Dalam penelitian ini akan diungkap bukti kesetaraan tersebut berdasarkan pada pengelompokan peserta lulusan Madrasah Aliyah Negeri – Sekolah Menengah Atas Negeri (MAN – SMAN, MS).

Pada umumnya tes potensi akademik (PA) dirancang untuk mengungkap kemampuan kognitif potensial, disusun berdasar konsep kemampuan dasar, memiliki kaitan minimal dengan kurikulum (Azwar, 2008). Pada SPMB-PTAIN 2012, PA dirancang terdiri dari enam komponen; yaitu (1) analogis, (2) logis, (3) analitis, (4) aritmetika, (5) komparasi, dan (6) geometri. Komponen-komponen

yang hendak diungkap ini tidak dapat dilepaskan secara penuh dengan silabus/kurikulum yang ada di sekolah. Materi komponen-komponen tersebut dipelajari di sekolah sehingga wajar apabila potensi akademik berkorelasi dengan prestasi akademik (Fraser, Walberg, Welch, & Hattie, 1987). Bahkan, Mehrens dan Lehmann (1991, p. 244) mengklaim bahwa tes potensi akademik mengukur *developed ability*, bukan *innate ability*, sehingga pada derajat tertentu kemampuan akademik yang ia miliki akan berpengaruh pula pada skor tes potensi. Artinya, kinerja peserta pada tes potensi akademik dipengaruhi oleh seluruh aktivitas yang telah ia lalui, termasuk ketika di sekolah. Dengan demikian, deteksi DIF berdasar kelompok sekolah relevan dengan dasar pengembangan PA SPMB-PTAIN.

Penelitian ini berupaya mengidentifikasi aitem-aitem PA SPMB-PTAIN yang mengandung DIF berdasarkan pada pengelompokan peserta lulusan Madrasah Aliyah Negeri – Sekolah Menengah Atas Negeri (MAN – SMAN, MS). Tujuannya adalah diperoleh informasi aitem-aitem PA mana saja yang mengandung DIF untuk selanjutnya dapat dilakukan perbaikan-perbaikan yang relevan dengan pencetus terjadinya DIF pada kelompok MS.

Differential Item Functioning

Berdasarkan pada *Standards for Educational and Psychological Testing* tahun 1999, keadilan dalam tes dapat dimaknai sebagai bebas bias pada level aitem ataupun tes (Penfield & Camilli, 2007). Evaluasi terhadap adil tidaknya aitem ini juga menjadi standar *Educational Testing Service* dalam mengembangkan produk-produknya (ETS, 2002). ETS menyebutkan bahwa secara empirik, prosedur mengevaluasi keadilan ini dilakukan dengan studi DIF.

Beberapa ahli memberikan batasan DIF, diantaranya Camilli (1992) menyebutkan bahwa sebuah aitem disebut berfungsi secara berbeda apabila probabilitas menjawab benar dipengaruhi oleh keanggotaan peserta pada kelompok. Angoff (1993) menyatakan bahwa istilah DIF mengacu pada perbedaan sifat-sifat statistik sebuah aitem antar kelompok dalam kemampuan yang sama. Dalam bahasa yang lain Hambleton dkk. (1991) menjelaskan bahwa sebuah aitem mengandung DIF jika individu-individu dengan kemampuan yang sama, tetapi berasal dari kelompok yang berbeda, tidak memiliki probabilitas menjawab benar

yang sama pada aitem tersebut. Penjelasan lain diberikan oleh Embretson dan Reise (2000) yang mengatakan bahwa DIF dikatakan terjadi saat sebuah aitem tes tidak memiliki hubungan yang sama dengan sebuah variabel laten, θ , antar dua atau lebih kelompok peserta tes.

Penfield dan Camilli (2007) mengemukakan bahwa DIF terjadi saat individu-individu dengan atribut laten yang besarnya sama namun berbeda kelompok, memiliki distribusi respons yang berbeda. Pada kasus aitem dikotomi, kelompok referensi dan fokal yang memiliki level θ sama akan memiliki probabilitas menjawab benar yang berbeda. Dengan demikian aitem berfungsi secara berbeda pada dua kelompok. Dalam bentuk persamaan, aitem yang tidak mengandung DIF akan mengikuti:

$$f(Y | \theta, G = R) = f(Y | \theta, G = F) \quad (1)$$

dimana:

G adalah variabel kelompok;

R mengacu pada kelompok referensi

F mengacu pada kelompok fokal

Y adalah probabilitas menjawab benar pada aitem

θ adalah atribut laten yang diukur oleh tes

Beberapa pendapat di atas mengarahkan pada pengertian bahwa sebuah aitem dikatakan teridentifikasi DIF jika terjadi probabilitas yang tidak sama dalam menjawab benar sebuah aitem pada dua kelompok peserta tes dengan kemampuan sama, setelah berada pada kontinum kemampuan yang sama.

METODOLOGI PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif berdasarkan data sekunder. Data ini diperoleh dari panitia SPMB-PTAIN tahun 2012. Total data peserta yang mengikuti SPMB-PTAIN pada tahun tersebut adalah 51570. Sebanyak 25744 mengerjakan pada paket 1, sementara sisanya mengerjakan

paket 2. Data yang dipergunakan dalam penelitian ini adalah data respons peserta pada paket 1.

Informasi banyaknya peserta yang mengerjakan soal paket 1 berdasarkan asal sekolah disajikan pada Tabel 1. Dengan pertimbangan jumlah peserta pada tiap sekolah asal dan kesetaraan status sekolah, peneliti merandom sebanyak 5000 peserta yang berasal dari MAN dan 5000 peserta dari SMAN.

Tabel 1. Rekapitulasi Peserta SPMB-PTAIN berdasarkan Asal Sekolah

Kategori	Jumlah Peserta	Jumlah Peserta (%)
MAN	5376	20,90
MAS	5308	20,60
PESA	502	1,90
SMEA	248	1,00
SMKN	1584	6,20
SMKS	938	3,60
SMAN	9148	35,50
SMUS	1869	7,30
LAIN	617	2,40
Kosong	154	0,60
Total	25744	100,00

Teknik analisis data yang digunakan dalam mengidentifikasi DIF dalam penelitian ini adalah *likelihood ratio test* (LRT) atau secara lengkap disebut *item response theory likelihood ratio test* (IRT-LRT). Pemilihan metode ini didasarkan pada pertimbangan:

- 1) Metode ini berbasiskan model, melibatkan parameter-parameter aitem secara langsung.
- 2) Metode ini paling sensitif bila dibandingkan dengan metode *Lord Chi-Square* dan *Signed and Unsign Area* dalam kerangka UIRT (Wagiran, 2005; Woods, 2009).

- 3) Kalibrasi dilakukan secara simultan terhadap parameter aitem pada kedua kelompok sehingga tidak perlu ada transformasi dari satu metrik ke metrik yang lain.
- 4) Kemudahan analisis sudah didukung oleh software IRTLRDIF yang dikembangkan oleh Thissen (2001).

HASIL PENELITIAN DAN PEMBAHASAN

Aitem yang dianalisis dipilih berdasarkan korelasi biserial (r_{bis}) dengan kriteria skor total. Sampel berdasarkan kelompok asal sekolah terdiri dari Madrasah Aliyah Negeri dan Sekolah Menengah Atas Negeri (MS) dengan ukuran 10000 peserta (5000 lulusan M dan 5000 lulusan S). Analisis berdasarkan kelompok ini memiliki hasil yaitu; 17 aitem selanjutnya tidak diikuti dalam analisis karena memiliki $r_{bis} < 0,25$, yaitu aitem-aitem dengan nomor urut: 3, 6, 7, 8, 10, 14, 16, 19, 22, 28, 46, 50, 64, 65, 66, 67, dan 68. Dengan demikian, analisis dilakukan terhadap 58 aitem.

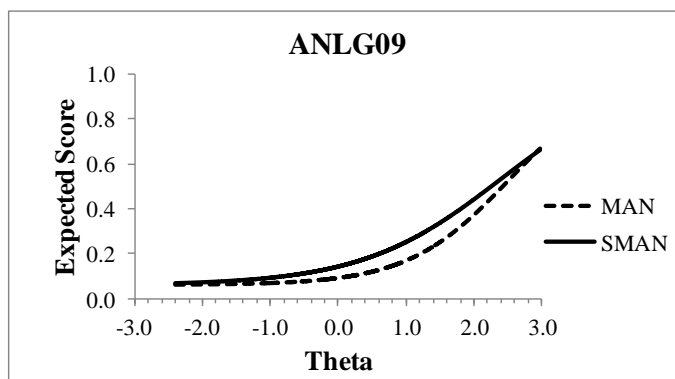
Berdasarkan 58 aitem yang diikuti dalam analisis, pemilihan model didasarkan pada kecocokan data dengan model yang digunakan. Informasi kecocokan yang disajikan pada Tabel 2 menunjukkan bahwa model 3PL lebih cocok dalam menjelaskan data respons para peserta tes, baik pada kelompok PL ataupun MS. Untuk itu analisis selanjutnya didasarkan pada model 3PL.

Tabel 2. Kecocokan banyaknya Parameter dalam Model

Model	-2 Loglikelihood	G ²	df	p	Keterangan
1PL	845621,96	-	-	-	-
2PL	831521,14	14100,8182	58	0,000	Model 2PL lebih informatif
3PL	784430,06	47091,0767	58	0,000	Model 3PL lebih informatif

Keterangan: 1PL = model logistik 1 parameter; 2PL = model logistik 2 parameter; 3PL = model logistik 3 parameter; G² = statistik G² (perbedaan likelihood); df = derajat kebebasan; p = probabilitas yang bersesuaian.

Analisis terhadap 58 aitem-aitem PA SPMB-PTAIN menunjukkan hasil dimana terdapat 15 aitem yang secara signifikan mengandung DIF. Sebagai ilustrasi, berikut ini ditampilkan beberapa *item characteristic curves* (ICC) aitem DIF yang menguntungkan lulusan SMAN (Gambar 1 dan Gambar 2) dan yang menguntungkan lulusan MAN (Gambar 3).



Gambar 1. ICC ANLG09 (menguntungkan lulusan SMAN)

Gambar 1 menyajikan ICC aitem ANLG09. Pada gambar tersebut nampak jelas bahwa pada sepanjang skala kemampuan peserta lulusan SMAN memiliki probabilitas menjawab benar yang lebih tinggi dibanding peserta lulusan MAN. Berarti dapat dikatakan bahwa aitem ini menguntungkan kelompok peserta lulusan SMAN.

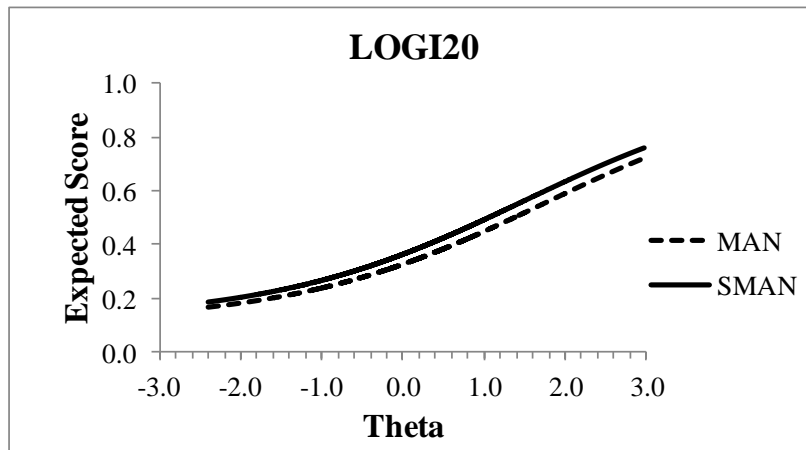
Redaksi aitem ANLG09 adalah sebagai berikut.

ABSOLUT

- (A) Liberal
- (B) Abstrak
- (C) Nisbi
- (D) Bebas
- (E) Labil

Untuk dapat menjawab kedekatan kata absolut pada aitem ANLG09, peserta tes perlu memahami makna kata absolut. Absolut memiliki makna tak terbatas, tak terkekang sehingga bebas berbuat apa saja. Dengan demikian jawaban yang tepat adalah (D) Bebas. Pola hubungan kedekatan makna kata dapat diselesaikan

melalui pemetaan hubungan antar kata dengan. Memori yang bekerja pada saat menyelesaikan aitem ini adalah *working memory* (WM) Halpern (2004).



Gambar 2. ICC LOGI20 (menguntungkan lulusan SMAN)

ICC LOGI20 yang disajikan pada Gambar 2 menunjukkan bahwa pada seluruh rentang θ kelompok lulusan SMAN memiliki probabilitas menjawab benar yang lebih tinggi dibandingkan peserta lulusan MAN. Oleh sebab itu dapat dikatakan bahwa lulusan SMAN diuntungkan pada aitem ini.

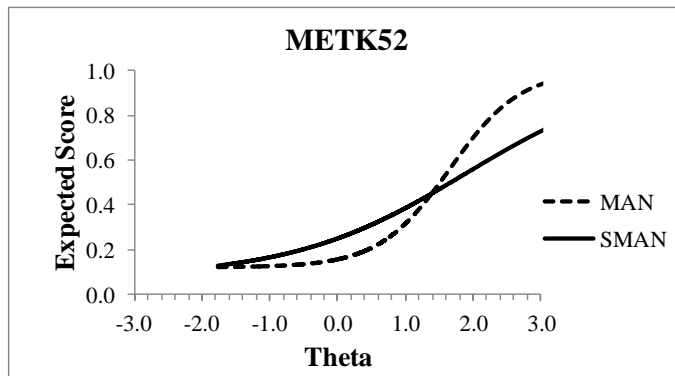
Redaksi aitem LOGI20 adalah sebagai berikut.

Semua Harimau adalah pemakan daging. Sebagian binatang adalah Harimau.
Jadi:

- (A) Semua pemakan daging adalah Harimau
- (B) Sebagian Harimau adalah pemakan daging
- (C) Sebagian binatang pemakan daging
- (D) a, b, c bukan jawaban yang benar
- (E) Tidak semua binatang pemakan daging

Guna menyelesaikan dengan benar aitem LOGI20, *working memory* (WM) sebagai bagian dari *fluid reasoning* (FR) memegang peranan penting. FR terkait dengan proses kognitif abstraksi penalaran yang melibatkan WM (Chuderski & Ngcka, 2012). Pada sisi lain, penyelesaian aitem tersebut bisa pula disandarkan pada pokok bahasan silogisme dalam pelajaran matematika. Berdasarkan pada kenyataan ini, dapat dikatakan bahwa ada kontribusi pembelajaran di sekolah

yang menyebabkan kelompok lulusan SMAN memiliki probabilitas menjawab benar yang lebih tinggi dibandingkan lulusan MAN.



Gambar 3. ICC METK52 (menguntungkan lulusan MAN)

Probabilitas menjawab benar aitem METK52 pada Gambar 3 menunjukkan bahwa pada kemampuan rendah sampai sedang ($-1,8 \leq \theta \leq 1,4$) aitem ini menguntungkan peserta lulusan SMAN. Sementara itu pada skala kemampuan tinggi peserta lulusan MAN lebih diuntungkan. Meskipun secara akumulatif aitem ini menguntungkan MAN, namun melihat rentang skala kemampuan kelompok SMAN diuntungkan pada skala kemampuan yang lebih lebar. Oleh sebab itu dapat dikatakan bahwa aitem ini hanya menguntungkan peserta lulusan MAN yang memiliki kemampuan tinggi.

Redaksi aitem METK52 adalah sebagai berikut.

Jika x adalah sisi persegi yang luasnya 25 cm^2 dan y adalah sisi panjang sebuah empat persegi panjang yang luasnya 50 cm^2 dengan sisi pendeknya 5 cm, berapakah xy ?

- (A) 25
- (B) 50
- (C) 20
- (D) 75
- (E) 55

Untuk dapat menyelesaikan aitem METK52 dibutuhkan kemampuan dasar persamaan luas persegi empat dan persegi panjang. Berdasarkan kemampuan

tersebut peserta akan dapat mengidentifikasi $x = 5$ dan $y = 10$. Dengan demikian jawaban yang benar adalah (B) 50.

Untuk menjawab benar aitem METK52 dengan cepat, strategi yang digunakan adalah strategi perhitungan umum yang diajarkan di sekolah melalui konsep luas persegi empat dan persegi panjang. Perhitungannya tidak dapat dilakukan melalui metode perkiraan secara cepat. Oleh sebab itu Gallagher, Levin, dan Cahalan (2002) menyebut aitem-aitem dengan strategi penyelesaian seperti ini sebagai aitem-aitem konvensional.

Pembahasan terhadap aitem-aitem yang terdeteksi DIF tidak dapat dituangkan secara keseluruhan dalam penelitian ini dikarenakan oleh batasan banyaknya halaman dalam tulisan ini. Oleh sebab itu, guna mempermudah uraian terhadap hasil analisis DIF dalam penelitian ini, berikut ini adalah rekap ringkasan DIF aitem-aitem PA SPMB-PTAIN tahun 2012 berdasarkan kelompok MS.

Tabel 3. Rekap DIF berdasarkan Kelompok MS

No	Aitem	UIRT	No	Aitem	UIRT
1	ANLG02	+	2	ANLG05	-
3	ANLG09	+	4	ANLG13	-
5	LOGI17	+			
6	LOGI18	+			
7	LOGI20	+			
8	METK39	+	12	METK52	-
9	METK40	+			
10	METK41	+			
11	METK45	+			
13	KOMP55	+			
14	KOMP56	+			
15	KOMP61	+			
			S	+	12
			M	-	3
			Jumlah		15

Berdasarkan informasi yang disajikan pada Tabel 3, penelitian ini menemukan sebagian besar aitem-aitem yang terdeteksi DIF terlihat menguntungkan kelompok peserta lulusan SMAN. Diantara 15 aitem DIF, 12 diantaranya menguntungkan peserta lulusan SMAN, 2 aitem sisanya menguntungkan peserta lulusan MAN.

Pada komponen analogis, 4 aitem DIF tersebar merata dengan 2 aitem menguntungkan lulusan SMAN dan 2 aitem menguntungkan lulusan MAN. Untuk komponen logis, 3 aitem menguntungkan lulusan SMAN. Pada komponen aritmetika, 4 aitem menguntungkan lulusan SMAN dan 1 aitem menguntungkan siswa lulusan MAN. Sementara itu pada komponen komparasi, 3 aitem menguntungkan siswa lulusan SMAN. Untuk komponen logis dan geometris, aitem-aitem tidak mengandung DIF berdasarkan kelompok MS. Hasil ini menunjukkan bahwa secara umum kelompok SMAN diuntungkan oleh aitem-aitem tes PA SPMB-PTAIN.

KESIMPULAN DAN REKOMENDASI

Kesimpulan yang dapat ditarik dalam penelitian ini adalah sebagai berikut.

- 1) ada 12 aitem menguntungkan siswa lulusan SMAN pada komponen analogi, logis, aritmetika, dan komparasi;
- 2) ada 3 aitem menguntungkan siswa lulusan MAN pada komponen analogi dan aritmetika;

Rekomendasi yang dapat diberikan adalah sebagai berikut.

- 1) Untuk tim pengembang dan penulis tes PA SPMB-PTAIN, perlu melakukan analisis isi secara lebih mendalam terhadap aitem-aitem yang terdeteksi DIF sehingga dapat dilakukan perbaikan-perbaikan pada pengembangan tes PA SPMB-PTAIN di masa mendatang agar meningkatkan validitas konstruk tes PA SPMB-PTAIN, juga menjamin kesetaraan dan keadilan bagi peserta lulusan MAN – SMAN;

DAFTAR PUSTAKA

- Angoff, W. H. (1993). Perspectives on Differential Item Functioning Methodology. Dalam W. H. Angoff (Ed.), *Differential Item Functioning* (Edisi ke-1, hh. 3-24). New Jersey: Lawrence Erlbaum Associates.
- Azwar, S. (2008). Kualitas Tes Potensi Akademik Versi 07A. *Jurnal Penelitian dan Evaluasi Pendidikan*, 12(2), 232-250.
- Camilli, G. (1992). A Conceptual Analysis of Differential Item Functioning in Terms of a Multidimensional Item Response Model. *Applied Psychological Measurement*, 16(2), 129-147.
- Chuderski, A., & Ngcka, E. (2012). The Contribution of Working Memory to Fluid Reasoning: Capacity, Control, or Both? *Journal of Experimental Psychology. Learning, Memory & Cognition*, 38(6), 1689-1710. doi: 10.1037/a0028465
- Embretson, S. E., & Reise, S. P. (2000). *Item Response Theory for Psychologist*. NJ: Lawrence Erlbaum Associates Inc.
- ETS. (2002). *ETS Standards for Quality and Fairness*. Princeton, NJ: Educational Testing Service.
- Fraser, B. J., Walberg, H. J., Welch, W. W., & Hattie, J. A. (1987). Syntheses of educational productivity research. *International Journal of Educational Research*, 11(2), 147-252.
- Gallagher, A. M., Levin, J., & Cahalan, C. (2002). Cognitive Patterns of Gender Differences on Mathematics Admissions Tests. *GRE Board Report No. 96-17P*. Princeton, NJ: Educational Testing Services.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. CA: Sage Publication Inc.
- Mehrens, W. A., & Lehmann, I. J. (1991). *Measurement and Evaluation in Education and Psychology*. Belmont, CA: Wadsworth/ThomsonLearning.
- Messick, S. J. (1998). Consequences of Test Interpretation and Use: The Fusion of Validity and Values in Psychological Assessment. *Research Report No. 98-48*. Princeton, NJ: Educational Testing Service.
- Penfield, R. D., & Camilli, G. (2007). Differential Item Functioning and Item Bias. Dalam C. R. Rao & S. Sinharay (Eds.), *Handbook of Statistics 26: Psychometrics* (Edisi ke-1, hh. 125-167). Amsterdam: Elsevier.
- Thissen, D. (2001). IRTLRDIF v.2.0b: Software for the computation of the statistics involved in item response theory likelihood-ratio tests for differential item

- functioning [Computer software]. Chapel Hill: L. L. Thurstone Psychometric Laboratory, University of North Carolina.
- Wagiran, B. K. (2005). *Perbandingan Berbagai Metode untuk Mendeteksi Bias Butir*. Yogyakarta: Program Pascasarjana UGM. Disertasi. Tidak diterbitkan.
- Woods, C. M. (2009). Empirical Selection of Anchors for Tests of Differential Item Functioning. *Applied Psychological Measurement*, 33(1), 42–57.